



The EU Mutual Learning Programme in Gender Equality


Artificial Intelligence and Gender Biases in Recruitment and Selection Processes

12-13 November 2020

Discussion paper



The information and views set out in this paper are those of the author(s) and do not necessarily reflect the official opinion of the Commission. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.



This publication is supported by the European Union Rights, Equality and Citizenship Programme (2014-2020).

This programme is implemented by the European Commission and shall contribute to the further development of an area where equality and the rights of persons, as enshrined in the Treaty, the Charter and international human rights conventions, are promoted and protected.

For more information see: http://ec.europa.eu/justice/grants1/programmes-2014-2020/rec/index_en.htm

Gender Biases in Selection and Recruitment Software

Miriam Kullmann

Harvard University, Weatherhead Center for International Affairs
WU Vienna University of Economics and Business

1. The use of AI in recruitment processes

The world of work is increasingly impacted by the use of ‘artificial intelligence’, that is machines or programmes that are being made to assume tasks usually carried out by humans, ranging, inter alia, from planning, learning, reasoning, problem-solving to knowledge representation.¹ Contemporary artificial intelligence is still largely narrow in a way, rather than general and encompassing a certain complexity, for it is directed at performing particularly defined tasks to solve ‘problems’, such as software that is built with the objective to help companies finding, but also keeping, the right employees.

Selection and recruitment software, which is part of human resources management (HRM), aims at easing the, at times, time-consuming recruitment processes. Recruitment software is often offered as cloud-based systems, replacing on-premise solutions based on a one-time licence, or as software-as-a-service (SaaS) model for which a monthly or annual subscription fee is paid.² The recruiting software landscape is quite diverse and, overall, aims at providing solutions for different problems, such as:³ CV parsing (i.e., replacing manual review of CVs); applicant tracking systems (i.e., electronic handling of recruitment and hiring needs); recruitment Candidate Relationship Management (i.e., engaging a company’s talent pool and see their work experience and their skills); and video interviewing (i.e., assessing job candidates and predicting an employee’s worth)⁴.

Overall, the benefit of such software is that it automates processes within an organisation (public and private), especially where employers experience a high recruitment and employee management volume, that is they can computerise and integrate several different HR management processes, including recruitment, training,

¹ N Heath, ‘What is AI? Everything you need to know about Artificial Intelligence’ (12 February 2020) ZDNet, www.zdnet.com/article/what-is-ai-everything-you-need-to-know-about-artificial-intelligence

² TechnologyAdvice Recruiting Software Buyer's Guide (updated 7 October 2020) technologyadvice.com/recruiting-software/#:~:text=Recruiting%20software%20is%20a%20category,for%20organizations%20to%20add%20employees

³ See for an overview: S Mondal, ‘The 38 Top Recruiting Software Tools Of 2020’ (1 April 2020) Ideal <https://ideal.com/top-recruiting-software>

⁴ A prominent example is HireVue, www.hirevue.com/platform/online-video-interviewing-software

payroll and administration of benefits, performance appraisal and analysis. In fact, traditional HR tasks are being ‘outsourced’ or ‘delegated’ to software, potentially saving costs and increasing efficiency and consistency on the selection and recruitment process. The underlying presumption is that the software is capable of matching skills and characteristics with the job demands in question.⁵ The use of HR management software falls in the broader context of what is referred to as workforce or people analytics⁶, meaning that data is being used to quantify and, based on that, analyse particular human traits, experiences, and skills of employees to make and justify hiring decisions, to promote and fire employees, thereby – seemingly – replacing, the biased ‘gut instinct’ or ‘anecdotal observation’.⁷

2. Potential sources of gender (and other) biases

The increased use of Artificial Intelligence in the world of work is cause for concern, for, as for instance the Advisory Committee on Equal Opportunities for Women and Men acknowledges, it can result in gender-based or other kinds of discrimination.⁸ As an increasing number of enterprises make use of selection and recruitment software, this may entail the risk of excluding, for instance, older job seekers and low-skilled unemployed individuals.⁹ A few examples will follow to illustrate the inherent risks involved in (semi-)automated decision-making processes.

A prominent example is that of Amazon’s machine learning algorithm that favoured male candidates because it had been trained on male-dominated resumes, which outnumbered female applications, over a period of ten years.¹⁰ Amazon had high expectations for its machine learning algorithm, namely through screening 100 resumes it should identify the five best candidates whom it would hire. Related to this example is that following an interesting study undertaken in the US, there are job-related gender stereotypes in online news sources. And as algorithms often emulate training datasets on which they are built, input that is biased will result in outcome that

⁵ MA Cherry, ‘People Analytics and Invisible Labour’ (2016) 61 *Saint Louis University Law Journal* 1, 1-2.

⁶ PT Kim, ‘Data-driven Discrimination at Work’ (2017) 58 *William and Mary Law Review* 857, 874.

⁷ Cherry, ‘People Analytics and Invisible Labour’ 1.

⁸ Advisory Committee on Equal Opportunities for Women and Men, Opinion on Artificial Intelligence – opportunities and challenges for gender equality (18 March 2020), https://ec.europa.eu/info/sites/info/files/aid_development_cooperation_fundamental_rights/opinion_artificial_intelligence_gender_equality_2020_en.pdf

⁹ European Parliament, *Employment and Skills Aspects of the Digital Single Market Strategy* (Directorate General for Internal Policies, Policy Department A, 2015), available at: https://digitalindustryalliance.eu/wp-content/uploads/2018/03/IPOL_STU2015569967_EN.pdf

¹⁰ InstightsTeam, ‘Overcoming AI’s Challenge In Hiring: Avoid Human Bias’ (29 November 2018) Forbes www.forbes.com/sites/instights-intelai/2018/11/29/overcoming-ais-challenge-in-hiring-avoid-human-bias/#2d5f6f4573bf

is biased.¹¹ Hence, the use of selection and recruitment software may entail the risk of replicating and furthering existing gender (and other) biases and thus contribute to what can be referred to as databased discrimination. Notably, algorithms use historical (i.e., past) data, some of which lack population diversity,¹² to make decisions for the future or even predict future behaviour. Algorithms may also be used to assess a candidate's personality, scrutinising also posts published on social media. For instance, Predictim rated a 24-year old woman as being at a very low risk concerning drug abuse, while being a higher risk for bullying, harassment and being disrespectful, without even providing an explanation for that decision. Such an assessment may change an initial positive opinion about a candidate.¹³

Against this background, the following statement by Iris Bohnet can be quoted. She writes that '[s]tereotypes serve as heuristics – rules of thumb – that allow us to process information more easily, but they are often inaccurate. What is worse, stereotypes describing how we believe the world to be often turn into prescriptions for what the world should be.'¹⁴ To understand where such 'biases' can come from, it is necessary to set out some general information on machine learning (algorithms). Algorithms are usually designed in a way that the decision leads to specified results or solving particular problems,¹⁵ that is finding the right employee for the job in question, with some values and interests preceding over others. They can, very simply, be defined as formally specified sequences of logical operations providing step-by-step instructions for computers to act on data and thus automate decisions.¹⁶ As follows from this definition, data forms the key input for algorithms to do their job. It is important to stress here that it is us, human beings, that produce the data and that data may be (consciously or not) be biased and may contain prejudices. It is also human beings that decide on what the aim of algorithms and the algorithmic model should be, that is what kind of problem should be solved with employing automated decision-making processes. It is, moreover, human beings that decide what kind of data is part of the training data set with which algorithms are being trained. In addition,

¹¹ C Chin, 'Assessing employer intent when AI hiring tools are biased' (13 December 2019) Brookings www.brookings.edu/research/assessing-employer-intent-when-ai-hiring-tools-are-biased/

¹² A Kaushal, R Altman and C Langlotz, 'Toward Fairness in Health Care Training Data' (October 2020) Stanford University Human-Centered Artificial Intelligence https://hai.stanford.edu/sites/default/files/2020-10/HAI_Healthcare_PolicyBrief_Oct20.pdf

¹³ D Harwell, 'Wanted: The 'perfect babysitter.' Must pass AI scan for respect and attitude' (23 November 2018) The Washington Post www.washingtonpost.com/technology/2018/11/16/wanted-perfect-babysitter-must-pass-ai-scan-respect-attitude

¹⁴ I Bohnet, *What Works: Gender Equality By Design* (Cambridge, Massachusetts, HUP 2016).

¹⁵ M Kullmann, 'Platform Work, Algorithmic Decision-Making, and EU Gender Equality Law' (2018) 34 *International Journal of Comparative Labour Law and Industrial Relations* 1, 9.

¹⁶ S Barocas and AD Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Review* 671, 674.

it is human beings that decide which software to use in an organization for particularly identified purposes.¹⁷

Technology, therefore, is not autonomous, but depends on human decision-making that is also not free of any innate or learned prejudices or biases.¹⁸ Furthermore, it is important to be aware of the fact that an algorithm can only be as good as the data it works with, meaning that if data is biased and that data is used by algorithms to automate decisions, their outcomes are most likely to be biased as well. There is therefore a high risk that algorithms continue to disadvantage historically disadvantaged groups if based on negative and unfounded assumptions.

There is still a crucial difference between human decision-making and algorithmic decision-making: an algorithm (usually) decides on the basis of predefined parameters and historical data. A human being, on the other hand, possesses freedom of choice as to whether the information provided through workforce analytics will determine outcomes in the decision-making process or whether, and to what extent, other information will also be taken into account. So far, intuition is a unique human characteristic.¹⁹ That does, however, not imply that humans are free of prejudices.

At the same time, algorithms can 'exploit the information in large datasets containing thousands of bits of information about individual attributes and behaviours', probably including information pertaining to a person's private sphere that might not be job- or work-related at all.²⁰ Not only can algorithms identify useful patterns in datasets, they also decide based on these patterns, usually much faster and thus more efficiently than a human.²¹ Above all, it may impact a much larger group of job candidates than human-based decision-making and thus may have wide-ranging societal implications if biases are reproduced on a structural basis. Overall, existing gender disparity in the workforce and biased datasets amplify gender inequality and project the potential injustice into the future.²² The latter has also been acknowledged by the EU's Gender

¹⁷ Hence, there are calls that Artificial Intelligence should always respect human agency and oversight and that humans should be in control at any time. See, e.g., European Parliament, Draft Report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)), available at: www.europarl.europa.eu/doceo/document/JURI-PR-650508_EN.pdf

¹⁸ F Stalder, 'Algorithmen, die wir brauchen' (2017) Netzpolitik.org <https://netzpolitik.org/2017/algorithmen-die-wir-brauchen/>

¹⁹ Kullmann, 'Platform Work' 12.

²⁰ *ibid*, with reference to Kim, 'Data-driven Discrimination at Work' 917ff. On privacy rights see: B van der Sloot, 'How to assess privacy violations in the age of Big Data? Analysing the three different tests developed by the ECtHR and adding for a fourth one' (2015) 24 *Information & Communications Technology Law*, 1.

²¹ Barocas and Selbst, 'Big Data's Disparate Impact' 677.

²² Kullmann, 'Platform Work'.

Equality Strategy 2020-2025,²³ following which women's employment levels in the EU are as high as ever, but they remain underrepresented in higher-paid professions and higher positions. Similarly, women are overrepresented in non-standard forms of work, often combining work with care duties. Such information may be reflected in the datasets used by recruitment software and therefore, as the example of Amazon shows, replicate societal perceptions and biases based on gender. It may therefore limit female workers access to particular jobs in cases where the recruitment software is trained with data that favours male workers.

3. Addressing potential risks of gender discrimination

Any policy that addresses the question how to counter potential risks of discrimination based on gender through Artificial Intelligence should, preferably, consider multi-pronged approaches. As biases are tenacious and difficult to eradicate in the short-term, it seems that there are a few issues that need in-depth consideration and discussion among a variety of stakeholders. In general, though, it should be noted at this stage that detecting databased discrimination may be difficult, even more so if compared with human decision-making. The following will provide a collection of five possible steps that may be useful and wary further discussion:

3.1 Raising awareness

Biases in the labour market²⁴ may have a huge impact on selection and recruitment decisions. Where decisions have become automated, even in part, these biases are very likely to be replicated. Hence, it is important to adopt a way through which stakeholders that are engaged with legal question and those that are involved in designing HR management software (the company that develops such software as well as the company that decides to use that software and sets certain requirements that must be met) should work more closely together to make explicit the legal and practical problems caused by biased outcomes. This particularly involves raising awareness of the specificity of each field to create a basis for a common understanding and for further exchange. While lawyers need to become more experienced with the software design and the processes that lead to the intended output, the 'technical' side should be open to the potential dangers involved in the software from a non-discrimination perspective, and in particular the protected characteristic of gender (and others as well). Mutual understanding and mutual learning in this area are arguably the first steps. Public engagement through other

²³ European Commission, A Union of Equality: Gender Equality Strategy 2020-2025, COM (2020) 152 final.

²⁴ On discrimination based on religion, see for instance, Case C-188/15 *Bouagnaoui* [2017] ECLI:EU:C:2017:204; Case C-157/15 *Samira Achbita* [2017] ECLI:EU:C:2017:203.

stakeholders, such as NGOs and equality bodies²⁵, could be another dimension of this awareness-raising, so as to strategically point out the implicit and explicit dangers created by biased databased discrimination.

3.2 Addressing legal issues

Combating discrimination and promoting gender equality can be seen as a fundamental social right, which is recognised in many EU and international as well as national legal systems. At the same time, it may not be ignored that, the principle of equal pay, laid down in Art 157 TFEU, which has been part of the EU since 1957, serves not only an economic aim (ie, preventing wage competition), but also a social aim. It is this dual role that is relevant in understanding the EU legal position in this context so as to consider potential solutions to reduce the risks of gender discrimination through Artificial Intelligence. In this context, reference can be made to the Commission Work Programme 2020, which contains a clear commitment to submit a proposal on pay transparency by the end of 2020.²⁶

The EU legal framework contains a few instruments that specifically address discrimination based on gender. Most recently, the European Pillar of Social Rights has gender equality as one of its key principles, addressing the issue of fair working conditions which also refers to access to employment.²⁷ That means that EU law protects individuals against discrimination based on sex regarding conditions for (initial) access to employment, to self-employment or occupation as well as the establishment, equipment or extension of a business or the launching or extension of any other form of self-employed activity (Directives 2006/54/EC and 2010/41/EC). Directive 2006/54/EC, in addition, expands the material scope so as to cover also discrimination based on sex with regard to working conditions including pay and occupational social security schemes (including those applicable to self-employed). It can be stated that, unless an algorithm leads to directly discriminatory outcomes, meaning that it decides based on the individual's gender whether or not to invite this person for a job interview, it seems that most such decisions would be classified as indirectly discriminatory, often using so-called proxies (eg, a women's age as being an indication of 'childbearing age') that are strongly linked to, and thus indicative of a worker's gender.

With indirect discrimination, however, the employer has the possibility to objectively justify its decision-making, the risk being that he might not know exactly how the selection and recruitment software came to take a certain decision. To assess a *prima facie* case of indirect discrimination, an individual needs to demonstrate, through significant statistical data, the adverse impact of the measure on persons

²⁵ See also R Allen QC and D Masters, *Regulating for an Equal AI: A New Role for Equality Bodies* (EQUINET, 2020), available at: https://equineteurope.org/wp-content/uploads/2020/06/ai_report_digital.pdf

²⁶ Commission Work Programme 2020: A Union that Strives for More, COM (2020) 37 final.

²⁷ Principle 2 in Chapter I on opportunities and access to the labour market.

characterised by gender as a protected ground. In general, the CJEU requires, at a very minimum, that the plaintiff shows a statistical disparity between, applied to our subject, the impact of the algorithmic process on the protected group and the comparator group. It follows from *Danfoss*, that plaintiff had to submit that average wage for women was lower.²⁸

However, the ability of the individual worker to have access to data that is related to such data runs against important obstacles, as it may be related to other (co-)workers, and as such are beyond the scope of data to which he or she can demand access. An interesting case in this context is that of *Kelly*, an unsuccessful candidate, where the CJEU denied the applicant any right to information in the possession of the organiser of that training on the qualifications of the other applicants in order to establish facts, which would lead to the presumption of direct or indirect discrimination.²⁹ Interestingly, the CJEU added that such a refusal may well affect the attainment of the objective pursued by the Directive and, in particular, may render this Directive ineffective in practice. This judgment constitutes a limitation for situations where there is a presumption of discrimination but it cannot be substantiated by sufficient facts.

Where decisions are based on large datasets, the question is whether it would be sufficient for a business to rely on decisions that do not necessarily reflect causal relations, but correlations. As a result, in cases in which the claimant can produce facts demonstrating that there has been discrimination, it should be the business that is required to establish whether the algorithm is valid if there is a suspicion that this is not the case; after all, it is the business that has – or at least should have – superior access to information about the design of the data model. This would also mean that the business would need to defend the accuracy of the correlations it would rely on to explain hiring decisions, meaning that it needs to show that the data or the algorithm are free of discrimination.³⁰ Moreover, it should not be possible for a business to invoke some kind of due diligence and then shift the risk and responsibility to the individual worker who has been discriminated as a result of biased HR management software decision-making. If the software is flawed, ie not in accordance with EU and national non-discrimination laws, then the employer should be liable and should have recourse to the company that has designed the software.³¹ Currently, a civil liability framework is considered at EU level, introducing regulation to ensure that those using ‘high-risk’ Artificial Intelligence can be held liable. Following the European Commission’s White Paper on Artificial Intelligence, a high-risk is expected to occur

²⁸ Case C-94/10 *Danfoss* [2011] ECLI:EU:C:2011:674.

²⁹ C-104/10 *Kelly* [2011] ECLI:EU:C:2011:506, paras 34, 38, 39.

³⁰ Kim, 'Data-driven Discrimination at Work' 921.

³¹ See the suggestion to introduce a civil liability framework at EU level for high-risk AI that causes damages: www.europarl.europa.eu/doceo/document/JURI-PR-650556_EN.pdf. There is some parallel here with the minimum wage liability in the context of the enforcement of the posting of workers’ legal regime, with the exception that the user company (the service recipient) has a way to invoke due diligence.

in particular sectors (e.g., healthcare, transport and parts of the public sector) and where the application is used in a way that high risks are likely to materialise. Automated decision-making in recruitment processes, as the White Paper stresses, impacts workers' rights, such as employment equality, and therefore is always considered to be a high-risk, meaning that recruitment and hiring software would need to satisfy certain requirements.³²

3.3 Assessing risks

Developing a framework based on which the potential risks of gender biases in HR management software can be assessed seems a valuable tool for the companies that design such software and those that decide to use the software. An interesting suggestion has been made by New Zealand, which adopted an Algorithm Charter in July 2020, in which it proposes a so-called risk matrix that helps to assess the likelihood (differentiating between: probably, occasional, improbably) of discrimination and its impact (distinguishing: low, moderate, high).³³ It is directed at improving government transparency and accountability, but could equally be helpful for private businesses. Overall, this Charter is, inter alia, aimed at striking a balance between privacy and transparency and preventing unintended bias.³⁴

³² Commission, 'White Paper On Artificial Intelligence - A European approach to excellence and trust' COM (2020) 65 final, 17-18. The requirements are (related to): training data; data and record-keeping; information to be provided; robustness and accuracy; human oversight; specific requirements for, e.g., remote biometric information. A follow-up on the White Paper will be published in Q1 2021.

³³ Algorithm charter for Aotearoa New Zealand (July 2020), available at: <https://data.govt.nz/use-data/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter>.

³⁴ As this Charter applies to New Zealand, it moreover aims to develop a Charter that reflects the principles of the Treaty of Waitangi.

Figure 1: Risk Matrix

Risk matrix

Likelihood

Probable Likely to occur often during standard operations			
Occasional Likely to occur some time during standard operations			
Improbable Unlikely but possible to occur during standard operations			
Impact	Low The impact of these decisions is isolated and/or their severity is not serious.	Moderate The impact of these decisions reaches a moderate amount of people and/or their severity is moderate.	High The impact of these decisions is widespread and/or their severity is serious.

This Risk Matrix is built on a few principles or values:

- Principle of transparency, requiring to clearly explain how decisions are informed by algorithms, which includes, inter alia, information on how data is collected, secured and stored.
- Role of people, stressing the need to actively engage with people and communities who have an interest in algorithms and, more importantly, consulting with those that are impacted by their use.
- Data should be fit for purpose, meaning that is important to understand the limitations of data and to identify and manage potential biases.
- Importance of human oversight, including the creation of a means for challenging or appealing to automated decisions and explain the role of humans in decisions that are informed by algorithms.

3.4 Auditing for algorithms

A further alternative might be found in systems of auditing for algorithms,³⁵ which could be transposed in the context of HR management software. Citron and Pasquale

³⁵ F Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, Massachusetts, Harvard University Press 2015) 150-151. See also FRA, *#BigData: Discrimination in data-supported decision making* (European Union Agency for Fundamental Rights, 2018), available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf

propose that (scoring) systems should be subject to licensing and audit requirements when they enter critical settings like employment, insurance, and health care.³⁶ That would mean that employers would have to disclose data related to outcomes to independent auditors or agencies. These independent actors would then be able to assess the potential biases of the algorithmic decisions themselves, as well as of the resulting consequences, and hence their suitability as a ground on which to base employers' choices. To provide an incentive to allow for the auditing to take place,³⁷ one might imagine a shift of the burden of proof for those employers basing their decisions on an un-audited rating system. Alternatively, to preserve the effectiveness of non-discrimination law, legislators should take steps to ensure that automated decisions are treated as an intrinsically insufficient ground for employment-related decisions.³⁸

3.5 Measuring outcomes

Another way that is relevant to discuss in this context is whether it is helpful that biased data and/or the algorithmic decision-making model could be changing and corrected. While cleaning the data seems to be cumbersome and problematic, as one may run behind for the data used may change frequently and the decision-making model as well, while the latter may have role in the following idea that can be suggested. A more fruitful possibility seems to be that of measuring outcomes, ie to compare the decisions that result from automated HR management software. Here, the different decisions will be compared, taking gender as one criterion, so as to see, in the case of a CV parsing software for instance, how many male/female candidates, ie the advantaged and disadvantaged group, submitted their application and how many of them were invited for a job interview or not. The idea would be to look at the decisions taken, which is similar to non-discrimination cases that have been brought before a civil or labour court. Here the judge usually will make an assessment based on statistical evidence and other facts that an individual provides to make believable that he/she has been discriminated. A judge then would be assisted by being provided with the outcomes/results of the decision-making process and can then decide which information he/she still needs to rule on the alleged discrimination. This would help with the highly contextualised non-discrimination cases that rely heavily on intuition and are open for judicial interpretation.³⁹

³⁶ DK Citron and F Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 *Washington Law Review* 1, 21-22.

³⁷ TZ Zarsky, 'Understanding Discrimination in the Scored Society' (2014) 89 *Washington Law Review* 1375, 1388.

³⁸ This section is based on R Ducato, M Kullmann and M Rocca, 'European Legal Perspectives on Customer Ratings and Discrimination' in T Abbabbo et al. (eds), *Performance Appraisal in Modern Employment Relations - An Interdisciplinary Approach* (London, Palgrave Macmillan 2019).

³⁹ S Wachter, BD Mittelstad and C Russell, 'Why Fairness Cannot be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI' (2020) SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547922